

Impact of a conventional introductory laboratory course on the understanding of measurement

Trevor S. Volkwyn, Saalih Allie,* and Andy Buffler

Department of Physics, University of Cape Town, Rondebosch 7701, South Africa

Fred Lubben

Department of Educational Studies, University of York, York YO10 5DD, United Kingdom

(Received 29 March 2007; published 1 May 2008)

Conventional physics laboratory courses generally include an emphasis on increasing students' ability to carry out data analysis according to scientific practice, in particular, those aspects that relate to measurement uncertainty. This study evaluates the efficacy of the conventional approach by analyzing the understanding of measurement of freshmen following the physics major sequence, i.e., top achievers, with regard to data collection, data processing, and data comparison, through pre- and postinstruction tests by using an established instrument. The findings show that the laboratory course improved the performance of the majority of students insofar as the more mechanical aspects of data collection and data processing were concerned. However, only about 20% of the cohort of physics majors exhibited a deeper understanding of measurement uncertainty required for data comparison.

DOI: [10.1103/PhysRevSTPER.4.010108](https://doi.org/10.1103/PhysRevSTPER.4.010108)

PACS number(s): 01.40.Fk

I. INTRODUCTION

Physics laboratory courses are typically centered around a set of tried-and-tested laboratory experiments¹ that illustrate various physics phenomena introduced in lectures. Students are expected to conduct their own experimental measurements and practice the standard procedures for dealing with these data.² A considerable body of research has focused on the learning effect of such "conventional" introductory laboratory courses on students who do not take physics as a major. The findings show that these students who have completed conventional laboratory courses only have a rudimentary understanding of the nature of scientific measurement and uncertainty. For example, after completing such a course, students were able to differentiate by rote between different types of uncertainties (or "errors") without showing an understanding of the implications for the quantity being measured (the measurand).^{3,4} Similarly, students were able to apply the routines for calculating means and standard deviations but were not able to interpret the spread in sets of repeated measurements.⁵⁻⁷ In another example, at the end of their laboratory course, students described measurements as "approximate" when it was felt that mistakes had occurred during the measurement process.⁸ Yet, the same students were able to apply rules of thumb for determining the uncertainty for a single measurement.

The understanding of the nature of scientific measurement has been given prominence in the descriptions of the goals of physics teaching by policy bodies such as the American Association of Physics Teachers.⁹ Consequently, the understanding of measurement has been included in the assessable outcomes of school science as reflected in international comparative studies^{10,11} and inventories of essential aspects of scientific literacy constructed by panels of experts.¹² We have developed a comprehensive research program into the understanding of the nature of scientific measurement and uncertainty of physics students.¹³⁻¹⁵ We have also looked at the effect of the explicit teaching of experimental tools and skills (as opposed to teaching for conceptual development)

and the use of "authentic contexts" for laboratory experiments.¹⁶ We reported that such teaching strategies considerably improved students' understanding of measurement, particularly in the areas of data collection and data processing, but less so for data comparison.¹⁴ These results reflected improvements from a very low base of understanding for South African nonphysics majors from educationally disadvantaged backgrounds taking an extended physics foundation program.¹⁶

In contrast to previous reports on students who were not typically physics majors, this paper focuses on students enrolled for the physics major course at the University of Cape Town, South Africa. We wish to address the assumption that the reported failure of fostering an understanding of measurement uncertainty may not apply for physics majors, i.e., students with the appropriate academic background and a keen interest in physics. A second difference with previous research work in this area is the fact that this study explores the effect of a conventional laboratory course rather than courses aimed at explicitly addressing issues related to measurement. Again, the assumption could well be that students who take the physics major sequence may not need such extra attention and will develop an appropriate understanding of uncertainty in a traditional laboratory course.

II. PURPOSE OF THIS STUDY

This study investigates the understanding of the nature of scientific measurement of a group of South African students following the physics major sequence, before and after a conventional introductory laboratory course, described below. In particular, the following questions guided the research project.

(1) What is the understanding of scientific measurement and uncertainty of the physics major students, before the laboratory course, in the areas of data collection, data processing, and data comparison?

(2) What are the changes in the understanding of the students in these three areas after completing the laboratory course?

III. CONVENTIONAL LABORATORY CURRICULUM

The introductory laboratory course at the University of Cape Town consisted of a series of 12 laboratory experiments, each of 3 h duration, that were spread over the academic year of 24 weeks. Each experiment was closely related to the theory that was being covered in lectures at the time. The majority of experiments involved either verifying various laws (e.g., Newton's second law and Boyle's law), measuring particular "constants" (e.g., freefall experiment to verify g and simple harmonic motion to measure the force constant ' k '), or learning to use and becoming familiar with laboratory apparatus (e.g., the multimeter and the oscilloscope). The students worked from a laboratory manual that contained instructions for carrying out each experiment, as well as descriptions of various apparatus, data analysis procedures, and conventions for writing up a laboratory report. Throughout the year, the students worked in pairs under the guidance of roving teaching assistants (TAs), who also graded the work by awarding an impression mark on a ten point scale at the end of each session. At the start of each laboratory afternoon, TAs gave students a short presentation in which they briefly outlined the main ideas of the experiment to be carried out and provided feedback from previous experiments. In addition, the laboratory course included two end-of-semester laboratory examinations in which an experiment (known to the students a week before the time) was individually carried out.

The first two weeks of the course were dedicated to the development of various experimental skills, including the use of measuring apparatus such as the Vernier calipers, micrometer screw gauge, counting apparatus, etc. The first two sessions also included lectures on experimental techniques, measurement, and uncertainty (following the approach as detailed in Taylor,² for example) and laboratory report writing. An exercise involving radioactive decay formed the basis for introducing ideas about random fluctuations and the normal (Gaussian) distribution. Rules of thumb were introduced to deal with specific measurement situations such as the "least-count" method of reporting the "error" for a single measurement. In particular, the idea of "significant figures" played a central role throughout. Uncertainty estimates were expected to accompany calculations based on measured data in all the experiments that were carried out.

IV. STUDENT COHORT

The particular cohort of students in this study may be generally characterized as having experienced good science teaching at school, including involvement in hands-on practical experiments. The analysis was based on responses from 53 students who participated in both pre- and postinstruction tests. As is frequently the case for physics major courses, the majority of the students (in this cohort, almost three quarters) were male.

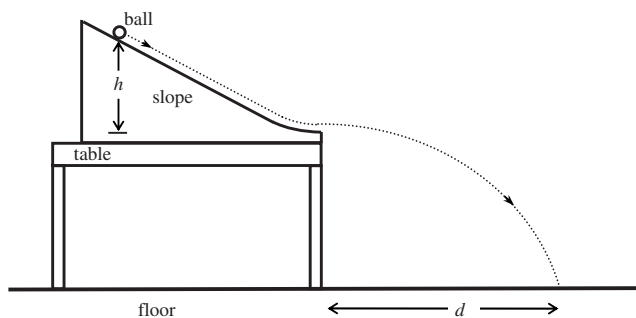


FIG. 1. The experimental context used for the probes.

V. METHODOLOGY

The research instrument before the course comprised a set of six written probes (questionnaire items). These items were deliberately taken to be identical to those used in previous studies^{5-7,13-15} since they had been independently validated¹⁷ for appropriate content and language and, most importantly, since this strategy allowed valid comparisons with outcomes of previous studies with nonphysics major students (see the Appendix for a brief explanation of the notation used to describe the probes as well as for the full text of all probes). Two probes (RD and RDA) dealt with the reasons for repeating measurements, thus covering the students' understanding of measurement in the area of data collection. Two probes (UR and SLG) surveyed students' ideas of measurement in the area of data processing, both numerically and graphically. The last two probes (SMDS and DMSS) dealt with measurement in the area of data comparison, focusing on the quality and comparability of data sets. The postinstruction probes were identical apart from the addition of two further data comparison probes (DMOS and DMSU). The DMOS probe is a variation of DMSS, while DMSU uses aspects of the formalism for comparing two sets of repeated measurements introduced in the laboratory course. Each probe presented a situation where a measurement decision was necessary and offered a number of alternative actions from which a choice was required. The reason for choosing a particular action was then requested in written form. Decisions that lead to actions are difficult to explore through written probes since respondents often have difficulty in visualizing "thought experiments." In order to minimize this problem, all the probes were related to the same experimental context (see Fig. 1). A large-scale version of the apparatus, a ramp with a horizontal edge and a ball, was also used to demonstrate the "experiment" before the probes were answered individually, in strict sequence and under examination conditions.

In the analysis, student responses were categorized according to the answer choice (A, B, or C) together with the different types of reasoning evidenced in the justification for this choice. The coding of the responses was undertaken expanding an alphanumeric scheme, which had been developed and validated previously.¹³⁻¹⁵ By using the expanded coding scheme, individual code assignment by the four researchers involved in the project yielded high levels of agreement (>90%). After discussion and refining of the coding

TABLE I. Students' reasoning when collecting data ($n=53$).

		Reasoning associated with paradigm after instruction (RD+RDA)		
		Consistent point reasoning	Consistent set reasoning	Total
Reasoning associated with paradigm before instruction (RD+RDA)	Consistent	1	6	7
	point reasoning	(2%)	(11%)	(13%)
	Mixed	0	7	7
	reasoning	(0%)	(13%)	(13%)
	Consistent	0	37	37
	set reasoning	(0%)	(70%)	(70%)
	Unclassifiable	0	2	2
	reasoning	(0%)	(4%)	(4%)
	Total	1	52	53
		(2%)	(98%)	(100%)

scheme, full agreement was achieved on assigned codes. Since the students in this cohort were largely first-language English speakers, they were expected to understand the probe questions and clearly express their views in written form.

Each response was classified according to whether or not the declared idea was compatible with the point paradigm or the set paradigm.¹⁴ The key notion of the point paradigm is that data are treated in a local realistic manner. Conclusions about the measurand directly follow from a selected datum that then represents the “true” value of the measurand. In general, a deviation from an expected result is seen as being due to environmental factors or experimenter mistakes. In cases where observations have been repeated, the dispersion is considered a result of the same influences: varying environmental influences or mistakes by the experimenter. In contrast, the key idea of the set paradigm is that each datum provides incomplete information about the measurand and that the best value of the measurand is obtained by combining the data at hand into a single (theoretical) quantity. The set paradigm is in line with the accepted view of scientific measurement, in which each reading is regarded as an approximation of the measurand. This view acknowledges that, in principle, knowledge about the measurand cannot be complete. All available readings are used together to construct the best approximation of the measurand together with an interval of uncertainty. Each student was then labeled according to one of three categories (consistent point reasoning, mixed reasoning, and consistent set reasoning) depending on whether their reasoning was consistent across the subset of probes relating to a specific area of measurement (data collection, data processing, or data set comparison).

VI. RESULTS

The results of the study are summarized in a series of tables below that show prepost frequencies based on the three categories: consistent point reasoning, mixed reasoning, and consistent set reasoning. Each category contains a wide range of responses, a small subset of which is illus-

trated by the quotes that lead up to each table. We present the results for the *sample* in the tables followed by a brief statistical analysis at the end of the section.

A. Students' ideas about the nature of scientific measurement when collecting data

Two probes (RD and RDA) explored students' ideas about measurement in the area of data collection. In one scenario, a group of students had collected one reading of the distance the ball moved from the table, and respondents were asked to justify if and why they needed to repeat the distance measurement (the RD probe). The RDA probe presented two different readings and students needed to justify if and why they needed to repeat the distance measurement again.

The two quotes below show examples of students who used point reasoning to guide their engagement with the task. “It’s useless to re-drop the ball from the same height if we already know that d is independent of h , because we dropped the ball under exactly the same conditions of speed (since we didn’t push the ball) and friction (we used same room).” (RDA response) “To release the ball a third time it would help to see which of the measurement is out when the ball is released for the third time.” (RDA response)

The two quotes below show reasoning that is consistent with the set paradigm but the justification for the stated action differs. The first quote articulates the idea of approaching a “true value” as the reason for collecting as many data as possible, while the second quote attributes physical effects to the spread. “If you take a number of results it is possible to find an average and therefore get a result closer to the true value.” (RD response) “One should always take as many measurements as possible and then calculate a mean value and associated standard deviation. In this experiment, it is particularly important to take many readings as the trajectory of the ball can be influenced by many things.” (RD response)

Table I summarizes the students' views of measurement for the two data collection probes.

The data in Table I show that 70% of the students in the sample provided RD responses for both probes that were com-

patible with the set paradigm prior to the laboratory course. In particular, students stated that more than one reading would be required in order to be able to calculate a mean. This included students who argued that calculating a mean would depend on whether or not there was scatter in the data. One in eight students (13%) consistently used point reasoning for both probes, while an equal fraction (one in eight) used reasoning associated with a different paradigm for each of the two probes, respectively. The use of both paradigms in this way is referred to as “mixed” in the tables. After instruction, almost all the students in the sample (98%) provided reasons that could be regarded as being consistent with the set paradigm. This is not surprising as the instructions that accompanied the experiments that were carried out throughout the year emphasized the repeating of measurements. It was notable that after completion of the course, nearly a third of the students explicitly linked repeating measurements to the “uncertainty” or “standard deviation” in the result, whereas fewer than 4% did so on entry. On the surface, therefore, these students appeared to recognize that the spread in a data set is an integral component of obtaining a measurement result.

B. Students’ ideas about the nature of scientific measurement when processing data

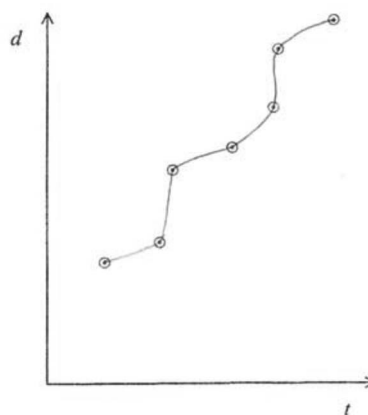
Student reasoning about measurement in the area of data processing was explored through two probes (UR and SLG). In the UR probe, a set of data was presented and students were asked what they would write down to represent the data set. The SLG probe required a trend line to be drawn on a set of covariant data.

Few responses for the UR probe indicated point reasoning. The quote below is a good example of student reasoning vacillating between the two different paradigms when decid-

ing on a representative value from a set of readings. “The 2nd release and 4th release agree or we can find a mean value by adding all the distances and dividing by 5.”

On the other hand, a student who appeared to have adopted set reasoning more completely stated the following: “This is the average of the releases. Although this exact result has not been observed, it is likely that after a large number of trials we would see that most results fall within a certain range of this “true” value.”

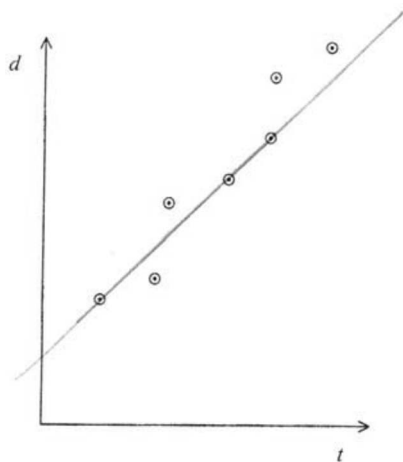
Some examples of how students dealt with representing a set of discrete data with an idealized curve are shown below. The first two examples are indicative of actions that follow from a point paradigm view, while the second two show different outcomes when performing the analysis from a set perspective.



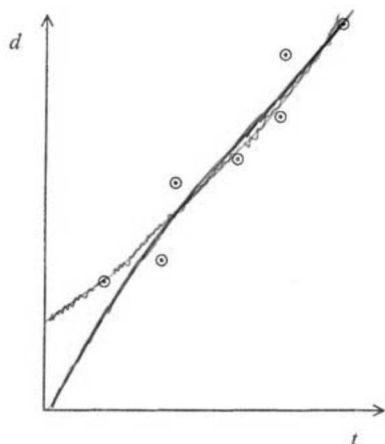
“All points I have joined by a flexible line which is not necessarily straight because the change in time from point to point is not the same as the distances are also not the same. This is because of the variation in the intervals of distance and also of time.”

TABLE II. Students’ reasoning when processing data ($n=53$).

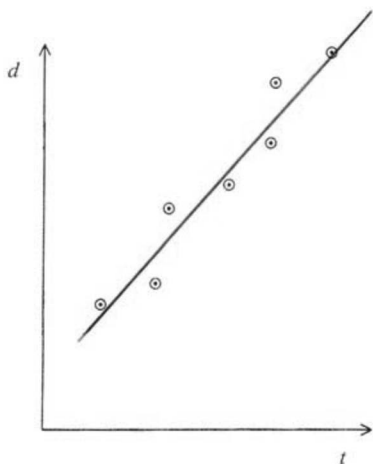
		Reasoning associated with paradigm after instruction (UR+SLG)		
		Mixed reasoning	Consistent set reasoning	Total
Reasoning associated with paradigm before instruction (UR+SLG)	Consistent	2	1	3
	point reasoning	(4%)	(2%)	(6%)
	Mixed reasoning	3	10	13
		(5%)	(19%)	(24%)
	Consistent set reasoning	0	37	37
		(0%)	(70%)	(70%)
	Total	5	48	53
		(9%)	(91%)	(100%)



“The graph that should be obtained is a straight line and the best thing to do is to join as many points as possible and in this case the maximum I can join to fit a straight line is three which I have joined.”



“The line has been drawn in such a manner that it encapsulates all the data and does not omit anything. The line has been drawn in relation to the general average result and thus does not pass through all the points. It is drawn in this way so that you can easily access the general average result in relation to the time.”



“The line drawn has been drawn in like an ‘average’ of all the points plotted, even though it does not go through any of the points, it is the line whose gradient will give the best representation of the results obtained.”

Table II displays the frequencies of the students’ reasoning, when asked to process data sets, numerically and graphically, in terms of how consistently their responses were commensurate with the point or set paradigms.

The data in Table II show that set reasoning substantially increased for both data processing probes after the completion of the laboratory course (91% of the sample). None of the students used only point reasoning after the laboratory course. Thus, after instruction, over three quarters of the students that used mixed reasoning on entry (19% of the sample) shifted to using set reasoning. It is interesting to note that five students (9% of the sample) continued to use point reasoning for generating a curve to a collection of covarying graphical data while calculating a mean for the set of repeated (numerical) readings.

C. Students’ ideas about the nature of scientific measurement when comparing data sets

Student views on measurement when comparing data sets were investigated by a series of probes that involved making a decision as to whether the data sets presented were in agreement or not (SMDS, DMSS, and DMOS). Two probes (SMDS and DMSS) were used prior to instruction, while an additional probe (DMOS) was used after the laboratory course. The SMDS probe involved a comparison of two sets of data, which had identical means but different spreads. For the DMSS probe, the data were such that the means were different but were spread over nearly identical intervals. In addition, the means of both sets of data fell within each other’s range. The third probe (DMOS) comprised data sets in which the means and the spreads differed but there was a partial overlap of the ranges. However, the mean of one data set fell outside the range of the other for this probe.

The quotes below show some of the ways that students reasoned about the agreement of data. These ranged from focusing on either the readings or the means being the same (point paradigm) to using the overlap of intervals to make the decision (set paradigm). “The two results don’t agree in that they have different answers. To agree, they should have exactly the same readings/answer.” (DMSS response) “The average for both groups differ hence they do not agree with the other, even though their average is in the same vicinity of the other group’s measurements.” (DMSS response) “Group A obtained answers between 422 and 440, B between 426 and 444. These ranges overlap. When the values for d are expressed as an interval, i.e., $\text{mean} \pm \text{standard deviation}$, the two groups’ answers will probably still coincide.” (DMSS response)

Table III shows the frequencies of students’ reasoning before and after instruction. Only students who consistently used set reasoning when answering all the data set comparison probes were classified as “consistent set reasoning.”

TABLE III. Students' reasoning when comparing data sets ($n=53$).

		Reasoning associated with paradigm after instruction (SMDS+DMSS+DMOS)			
		Consistent point reasoning	Mixed reasoning	Consistent set reasoning	Total
Reasoning associated with paradigm before instruction (SMDS+DMSS)	Consistent	8	6	4	18
	point reasoning	(15%)	(11%)	(8%)	(34%)
	Mixed	1	28	6	35
	reasoning	(2%)	(53%)	(11%)	(66%)
	Consistent	0	0	0	0
	set reasoning	(0%)	(0%)	(0%)	(0%)
	Total	9	34	10	53
		(17%)	(64%)	(19%)	(100%)

Prior to instruction, about a third of the students in the sample used point reasoning for both the SMDS and DMSS probes. The remaining two-thirds used reasoning associated with the set paradigm for only one of the two probes (mixed reasoning in Table III). Closer inspection revealed that almost all of these students recognized that the degree of spread in a data set is indicative of measurement quality, as demonstrated in their responses to the SMDS probe but focused only on the individual readings or the means of the data sets when deciding whether two data sets agreed for the DMSS probe. More than 80% of the students who used mixed reasoning prior to instruction continued to do so after instruction. Of the 34 students who were categorized as using consistent point reasoning before the course, only 8 (24% of this group) students ended up in the consistent set reasoning category. In summary, only ten students (19% of the sample) consistently used reasoning associated with the set paradigm after the laboratory course. These data also show that the shift observed for students adopting consistent set reasoning was small when compared to the shifts indicated for data collection and data processing.

One of the observations that has been made in previous studies has been that students appear to be able to answer questions that are more “formally” posed but that these responses cannot be used to gauge the degree of understanding of the concepts involved. Table IV contrasts individual stu-

dents' reasoning for answering a “formal” question where numerical values for a mean and standard deviation are presented (probe DMSU), compared with a student classification based on their reasoning in all three areas, i.e., data collection, data processing, and data set comparison (as described earlier). Students were classified as consistent set reasoners if their responses across probes were consistent with the set paradigm, in each of the three areas of measurement. Thus, Table IV can be interpreted as showing the relationship between the algorithmic aspects of the set reasoning (“formal probe”) versus an understanding of the set paradigm (student classification).

From Table IV, it is clear that even though 58% of the students in the sample used set reasoning when presented with data in the formal manner, less than a third of this group (17% of the sample) were consistently classified as reasoning in terms of the set paradigm. This indicates that students who are able to use the formalism correctly do not necessarily have a deep understanding of the nature of uncertainty.

The results shown in the tables are exact for the sample in question. We present a brief statistical analysis, for the purpose of further generalization, by comparing the proportions of students who consistently used set reasoning before and after instruction (Tables I–III) and set reasoning after instruction (Table IV). These proportions are shown in Table V below together with their 95% confidence intervals. Figure 2

TABLE IV. Students' reasoning after instruction ($n=53$).

		Reasoning associated with paradigm after instruction (DMSU)		
		Mixed reasoning	Consistent set reasoning	Total
Reasoning associated with paradigm after instruction (RD+RDA+UR +SLG+SMDS +DMSS+DMOS)	Mixed reasoning	21 (40%)	22 (41%)	43 (81%)
	Consistent set reasoning	1 (2%)	9 (17%)	10 (19%)
	Total	22 (42%)	31 (58%)	53 (100%)

TABLE V. Statistical analysis of proportions of students using set reasoning consistently. The numbers in parentheses define 95% confidence intervals.

Table	Percentage of students using set reasoning consistently		McNemar's test p value
	Before instruction	After instruction	
I	70 (56, 82)	98 (90, 99.95)	0.00006
II	70 (56, 82)	91 (79, 97)	0.001
III	0 (0, 7)	19 (9, 32)	0.002
	After instruction (all probes except DMSU)	After instruction (DMSU)	
IV	19 (9, 32)	58 (44, 72)	0.000006

summarizes the data in graphical form. We also formally tested whether the proportions significantly differed. This calculation has to take into account that the probes are applied to the same group of students before and after instruction and, hence, the results are not independent. Thus, McNemar's test¹⁸ for matched binomial proportions was used as an appropriate procedure. The resulting p values are indicated in Table V.

The analysis shows that in cases I–III, comparing pre- and postresponses, the observed differences between the proportions are significant ($p=0.000\,06$, 0.001 , and 0.002 , respectively). However, only in cases I and II, where the preinstruction proportions are already high, would the improvements be regarded as pedagogically acceptable. In case III, the pedagogical outcome is not successful given the actual postinstruction proportions spanned by the 95% confidence interval in question (9%–32%). In case IV, the observed difference in the proportion of students who were able to apply the rules of the formalism (for comparing data sets), compared with the proportion classified as consistent set reasoners (on the basis of all the previous probes), is significant ($p=0.000\,006$).

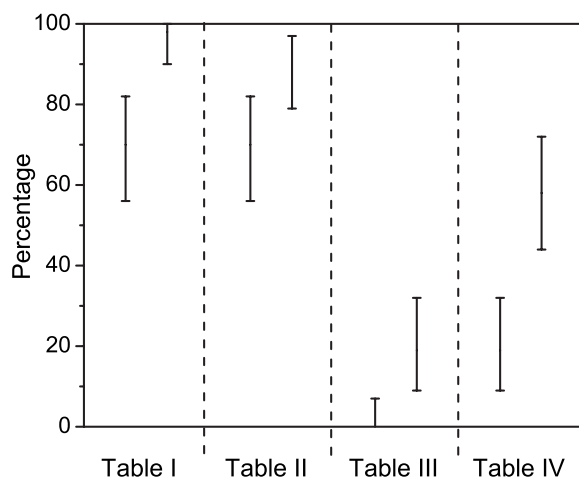


FIG. 2. Proportions of students in Tables I–IV using consistent set reasoning. For Tables I–III, the bars on the left indicate the collected preinstruction data, and the bars on the right the postinstruction data. For Table IV, the left bar (overall categorization) and the right bar (formal probe only) are both postinstruction data. The bars indicate 95% confidence intervals.

VII. DISCUSSION AND CONCLUSIONS

A. Summary of results from this study

We have explored the understanding of measurement held by prospective physics majors. Before any instruction, our observations showed that two out of three students provided responses, which were associated with the set paradigm in probes dealing with collecting and processing measurement data. However, none of these students consistently used set reasoning when comparing data sets. A conventional laboratory course was successful in consolidating and improving the students' understanding of measurement for data collection and data processing: virtually, all students in the sample consistently used set reasoning after the course in these two areas, an improvement which is statistically and pedagogically significant. However, after this course, only 19% of the sample provided responses associated with the set paradigm to probes dealing with data comparison, an increase which is statistically significant but not pedagogically acceptable. The conclusion may therefore be drawn that the strategies used in the conventional introductory laboratory course were unsuccessful in improving the students' understanding of uncertainty beyond the appropriation of the numerical routines.

This conclusion is borne out by the emergence of overall “mixed reasoners.” In the probes dealing with data collection and data processing, questions dealing with procedures such as repeating the measurement several times, calculating a mean, etc., are easily answered by rote, i.e., students in the mixed reasoning group were using the surface features¹⁹ of the set paradigm. Thus, rather than having made sense of the data analysis framework as a whole, they have learnt to strategically use certain algorithms. This is consistent with the observation that more than half of the students who used mixed reasoning after instruction explicitly referred to the formal constructs of standard deviation and/or uncertainty in their written responses. Thus, if we were to rely on the traditional forms of assessment, we would conclude that about three-fifths of the students had mastered the data analysis aspects of the laboratory course. However, our findings show that a very much smaller proportion of the students (one-fifth) demonstrated a coherent understanding of the nature of measurement uncertainty.

Formulating student understanding of scientific measurement in terms of the point and set paradigms has been shown to be a useful tool in a number of research programs to

date.^{5-7,13-15} We wish, however, to emphasize that the “point” categorization of a student is not meant to be pejorative, nor should it be used to identify “student misconceptions” regarding measurement. Our view is that student responses are framed by the perceived task and that “pieces of knowledge”²⁰ are activated by the context that is presented. In turn, the way in which students frame a task depends on previous experience and on whether or not prior learning has allowed students to make sense of the situation or whether they have the appropriate mental models in place.

B. Comparison with results from other studies

The results from the present study may be compared with a previous study¹⁴ at the University of Cape Town with a group of nonphysics majors. Since the latter typically had little or no prior laboratory experience from school, the introductory physics laboratory course,¹⁶ which these students completed, therefore contained additional activities that were designed to expose the students to apparatus and develop basic measurement skills. However, the course dealt with measurement uncertainty in the same way as the course described in the present study, and both cohorts of students completed similar data analysis tasks. It was found that about 40% of the nonphysics majors gave responses classified as being consistent with the set paradigm after instruction in the areas of data collection and data processing, with another 40% being classified as mixed reasoners. This contrasts with the present study in which over 90% of the physics majors were classified as set reasoners in these two areas. However, the proportions of students, whose postinstruction reasoning in the area of data set comparison were classified according to the set paradigm, were found to be nearly identical in the present study (19%) and the previous study (21%). This suggests that the assumption that students with high levels of interest in physics and adequate schooling backgrounds do better in developing an appropriate understanding of measurement uncertainty appears to be unfounded.

Recently, a number of studies into the effect of innovative laboratory curricula on students’ understanding of experimentation and measurement have been reported.^{5-7,21-23} The Scientific Community Laboratories (SCL)^{5,21} are concept-based laboratory activities developed at the University of Maryland, with components structured around increasingly complex concepts of measurement. The course starts with an open-ended exercise aimed at strengthening students’ ability to distinguish between descriptive tasks (summarizing data) and predictive tasks (using the data to extrapolate). Subsequent exercises explicitly address the purpose of multiple measurements, the use of range overlap, and systematic versus random mechanisms. The latter laboratory exercises help students to estimate uncertainties for a variety of experiments and to develop ways of thinking about discrepancies between measured results and theory.

The Investigative Science Learning Environment (ISLE)^{22,23} at Rutgers University provides students with laboratory experiences aimed at improving their “process skills.” Several of these skills, which reflect scientific practice, depend on a good understanding of measurement and

uncertainty. The ISLE program is organized around three types of experiments. Observational experiments require students to investigate new phenomena, collect and analyze their own data, and look for patterns leading to the formulation of hypotheses. In testing experiments, students make predictions about the outcomes of experimental investigations based on hypotheses derived from their physics knowledge and then design experimental procedures to test the hypotheses. Application experiments require students to solve authentic problems and determine unknown quantities. In all three experiment types, an understanding of the quality of the data collected and the experimental procedures used is emphasized.

The effectiveness of the teaching strategies used with nonphysics majors in these laboratory courses has been reported.^{21,23} The evaluation of the course at the University of Maryland²¹ indicated increases in the number of students that exhibit an understanding of measurement uncertainty but this amounted to fewer than half of the participating students. The evaluation of the ISLE teaching strategies at Rutgers University²³ showed that changes in students’ ability to evaluate experimental uncertainties were not significant. Thus, in general, the results of these studies together with those of others^{6,7} suggest that students leave introductory physics laboratory courses with an ability to carry out certain procedures but lack a coherent understanding of the nature of uncertainty, irrespective of the student profile, or the way in which the laboratory course was delivered. This raises the question as to why students arrive at a university with a view of scientific measurement that is typically described by the point paradigm and which is not significantly altered by present forms of the introductory physics laboratory course.

C. Consequences and recommendations for course design

The experience of measurement is rooted in everyday life where the terms approximate and “exact” are often used loosely and related to the notion of “good enough for this purpose.” The precision required in everyday measurement strongly depends on the context. For example, weighing out flour when baking a cake is not as critical as measuring the dose of a baby’s medication. Hence, it is likely that a student’s view of the nature of science as an exact enterprise is responsible for a view of the nature of scientific measurement as the pursuit of exact results. Our findings in this area²⁴ have suggested that a student’s view of the nature of science does, indeed, have a bearing on his or her view of the nature of scientific measurement. In support of this view, Séré *et al.*²⁵ presented students with measurements in different science disciplines (biology and physics) and in everyday situations. They found that students used different epistemologies and ontologies of the nature of science for processing the data in the various contexts. The notion of an epistemology of the *nature of measurement*, as distinct from an epistemology of the nature of science, is introduced. Before instruction, many students believe that uncertainty in measurement can be reduced to zero. Furthermore, even in situations where conceptual understanding of measurement and uncertainty has been strongly emphasized in the curriculum,

it has been repeatedly found that, apart from the few remaining consistent point reasoners, most students display competence in the numerical tools of data analysis, but few are able to conceptually reason.

We have previously argued²⁶ that one of the key stumbling blocks in the path of understanding is the statistical formalism of data analysis used in most introductory laboratory courses that relies on analyzing data in terms of frequencies (and is hence often termed “frequentist”²). In contrast, the probabilistic interpretation of measurement as advocated by the international bodies for metrology^{27,28} results in a framework in which the interpretation of uncertainty is clearer and more tangible and provides a coherent way for evaluating uncertainties of single and multiple measurements.²⁹ In addition, the scheme equally applies when treating different sources of uncertainty, obviating the need for *ad hoc* prescriptions and rules of thumb.²⁶ In the frequentist framework, it is usually assumed that observations of the measurand are randomly scattered around a “true” value. Thus, the true value of the measurand is considered to have no uncertainty associated with it: the data themselves are regarded as being “uncertain.” In contrast, in the probabilistic framework, the data are regarded as the manifestations of the phenomenon and are treated as constants, while it is the inference that is made about the measurand, which has a degree of uncertainty associated with it. In other words, what we may conclude about the measurand must necessarily be incomplete since our knowledge about the measurand is always based on a finite data set. The measure of the degree of incompleteness of information about the measurand is provided by the uncertainty parameter.

We have designed and implemented a course³⁰ based on the probabilistic framework of metrology, which provides opportunities for students to explore the nature of uncertainty in measurement through activities that challenge notions of measurement yielding an exact (pointlike) result. For example, students are asked to read both analog and digital scales of increasing sensitivity. Reading analog scales always requires judgment on the part of the observer and hence results in imperfect knowledge. A digital instrument can always be designed to display further digits in principle but not in practice; hence, the inference from a digital display is an interval centered around the digit displayed. In both cases, students come to realize that even in the absence of all other forms of uncertainty, the *inference* that follows from reading an instrument will always result in an interval, not a point. The course starts with students having to list all the factors that they think might have influenced their measurements and qualitatively estimate the effect of these factors. This provides opportunities to consider these factors and to discuss how these influences can be modeled, which culminates in the numerical description of uncertainty. One of the features of the probabilistic approach is that it provides a consistent framework for estimating uncertainties for both single and repeated observations where the data are dispersed. Finally, the measurement result is interpreted as a statement of the available *knowledge* or information about the measurand,

an approach which we believe provides persuasive pedagogic opportunities. We have avoided the term error, as suggested by the international bodies for metrology,^{27,28} but more importantly, we have tried to have students engage with the concepts and ideas before attaching terms to them.

D. Conclusion

In conclusion, it is clear that at this introductory physics level, an appropriate understanding of scientific *measurement* depends critically on an appropriate understanding of the nature of *uncertainty* in measurement. Conversely, it may be argued that the introductory physics laboratory course should be designed so as to explicitly develop a conceptual understanding of uncertainty and not only teach the mathematical calculations used to quantify uncertainty. However, the results of the present work, and others, suggest that innovative laboratory curricula that provide a range of hands-on activities that expose the nature of measurement uncertainty will only be partially successful in developing an appropriate understanding of uncertainty, which underpins scientific measurement. The reporting of scientific measurement, as a form of scientific evidence, requires that the degree of the knowledge be communicated (in the form of a numerical uncertainty) in a consistent way. The conceptual underpinnings that allow these numerical estimates of uncertainties to be generated should therefore also form part of the introductory physics laboratory. We suggest that the statistical formalism that forms part of most laboratory curricula be examined for logical and pedagogical consistencies and that the probabilistic framework be considered as a viable alternative. As with any other aspect of physics teaching and learning, it is possible for students to navigate their way through their laboratory course by relying on surface (rote) learning of procedures unless students have opportunities to discover that their mental models of scientific measurement may not necessarily be appropriate in the scientific context. A recent evaluation³¹ of our course based on the probabilistic framework has shown that we have been significantly more successful in the development of an appropriate conceptual framework of uncertainty in scientific measurement.

ACKNOWLEDGMENTS

We thank the University of Cape Town and the National Research Foundation (South Africa) for financial support in carrying out this work. We also thank Mary Lou Thompson for her advice regarding the statistical analysis.

APPENDIX: PHYSICS MEASUREMENT QUESTIONNAIRE

The following shorthand notation is used to describe the probes in the text: RD (repeating distance measurements); RDA (repeating distance again); UR (using repeated measurements); SLG (straight line graph); SMDS (same means different spread); DMSS (different means same spread); DMOS (different means overlapping spread); DMSU (different means same uncertainty).

*saalih.allie@uct.ac.za

- ¹F. Tyler, *A Laboratory Manual of Physics*, 6th ed. (Arnold, London, 1988).
- ²J. R. Taylor, *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*, 2nd ed. (University Science Books, New York, 1997).
- ³M.-G. Séré, R. Journeaux, and C. Larcher, Learning the statistical analysis of measurement error, *Int. J. Sci. Educ.* **15**, 427 (1993).
- ⁴J. Garatt, A. Horn, and J. Tomlinson, *Misconceptions about error*, *Chemistry Education Research and Practice* **4**, 54 (2000).
- ⁵R. Lippmann, Ph.D. thesis, University of Maryland, 2003.
- ⁶D. Deardorff, Ph.D. thesis, North Carolina State University, 2001.
- ⁷D. Abbott, Ph.D. thesis, North Carolina State University, 2003.
- ⁸D. Evangelinos, E. Psillos, and O. Valassiades, in *Teaching and Learning in the Science Laboratory*, edited by D. Psillos and H. Niederrerr (Kluwer Academic, Dordrecht, 2002), pp. 179–190.
- ⁹American Association of Physics Teachers, Goals of the introductory physics laboratory, *Am. J. Phys.* **66**, 483 (1998).
- ¹⁰Particularly in surveys of scientific literacy through the Programme for International Student Assessment (PISA) as in OECD (Organisation for Economic Cooperation and Development), *PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills* (OECD, Paris, 2003).
- ¹¹M. Lemke and P. Gonzales, *U.S. Student and Adult Performance on International Assessments of Educational Achievement: Findings from The Condition of Education 2006* (U.S. Department of Education, Washington, DC, 2006).
- ¹²J. Osborne, S. Collins, M. Ratcliffe, and R. Duschl, What ‘ideas-about-evidence’ should be taught in school science? A Delphi study of the expert community, *J. Res. Sci. Teach.* **40**, 692 (2003).
- ¹³S. Allie, A. Buffler, L. Kaunda, B. Campbell, and F. Lubben, First-year physics students’ perceptions of the quality of experimental measurements, *Int. J. Sci. Educ.* **20**, 447 (1998).
- ¹⁴A. Buffler, S. Allie, F. Lubben, and B. Campbell, The development of first year physics students’ ideas about measurement in terms of point and set paradigms, *Int. J. Sci. Educ.* **23**, 1137 (2001).
- ¹⁵F. Lubben, A. Buffler, S. Allie, and B. Campbell, Point and set reasoning in practical science measurement by entrant university freshmen, *Sci. Educ.* **85**, 311 (2001).
- ¹⁶S. Allie and A. Buffler, A course in tools and procedures for Physics I, *Am. J. Phys.* **66**, 613 (1998).
- ¹⁷As replicated in E. Redish, *Teaching Physics with the Physics Suite* (Wiley, Hoboken, NJ, 2003).
- ¹⁸H. Motulsky, *Intuitive Biostatistics* (Oxford University Press, New York, 1995).
- ¹⁹F. Marton and R. Säljö, On qualitative differences in learning: 1. Outcome and process, *Br. J. Educ. Psychol.* **46**, 4 (1976); On qualitative differences in learning: 2. Outcome as a function of learners’ conception of task, **46**, 115 (1976).
- ²⁰R. E. Scherr, Modeling student thinking: An example from special relativity, *Am. J. Phys.* **75**, 272 (2007).
- ²¹R. L. Kung, Teaching the concepts of measurement: An example of a concept-based laboratory course, *Am. J. Phys.* **73**, 771 (2005).
- ²²E. Etkina, A. Van Heuvelen, D. T. Brookes, and D. Mills, Role of experiments in physics instruction—a process approach, *Phys. Teach.* **40**, 351 (2002).
- ²³E. Etkina, S. Murthy, and X. Zou, Using introductory labs to engage students in experimental design, *Am. J. Phys.* **74**, 979 (2006).
- ²⁴B. Ibrahim, M.Sc. thesis, University of Cape Town, 2006.
- ²⁵M.-G. Séré, F. Fernandez-Gonzalez, J. Gallegos, F. Gonzales-Garcia, E. De Manuel, J. Perales, and J. Leach, Images of science linked to labwork: A survey of secondary school and university students, *Res. Sci. Educ.* **31**, 499 (2001).
- ²⁶S. Allie, A. Buffler, B. Campbell, F. Lubben, D. Evangelinos, D. Psillos, and O. Valassiades, Teaching measurement in the introductory physics laboratory, *Phys. Teach.* **41**, 394 (2003).
- ²⁷International Organization for Standardization, *International Vocabulary of Basic and General Terms in Metrology (VIM)* (ISO, Geneva, 1996).
- ²⁸International Organization for Standardization, *Guide to the Expression of Uncertainty in Measurement (GUM)* (ISO, Geneva, 1995).
- ²⁹L. Kirkup, A guide to GUM, *Eur. J. Phys.* **23**, 483 (2002).
- ³⁰A. Buffler, S. Allie, and F. Lubben, *Teaching Measurement and Uncertainty the GUM Way*, *Phys. Teach.* (to be published).
- ³¹S. Pillay, A. Buffler, F. Lubben, and S. Allie, *Effectiveness of a GUM-Compliant Course for Teaching Measurement in the Introductory Physics Laboratory*, *Eur. J. Phys.* (to be published).